

“CLUSTERING”

Artículo

MARTHA CAROLINA OSPINA RODRÍGUEZ

Código: 33042218

HENRY ALEXANDER RENGIFO SÁNCHEZ

Docente

UNIVERSIDAD DE LA SALLE

FACULTAD DE SISTEMAS DE INFORMACIÓN Y DOCUMENTACIÓN

RECUPERACIÓN DE INFORMACIÓN

BOGOTÁ D.C.

2007

“CLUSTERING: GESTIÓN DE CONOCIMIENTO EMPRESARIAL”

Resumen:

Este artículo inicia con una introducción que sitúa al lector en el marco de las tecnologías de información, su relevancia e influencia en los procesos empresariales a nivel interno y su relación con la sociedad del conocimiento, la economía y las tecnologías de información. A continuación se explica brevemente el significado de los clúster, sus características, los métodos estadísticos de agrupación y utilidad dentro de la organización. Se plantea el algoritmo ISODATA como una de las opciones más efectivas para la clasificación y agrupación de los clúster evidenciando su aplicabilidad en la gestión de información y conocimiento de una empresa a través de cooperaciones y alianzas entre la investigación, la ciencia, el desarrollo tecnológico, la innovación y el rendimiento desde todas sus perspectivas que ofrecen una neo visión de las dimensiones organizacionales.

Palabras Clave:

Clúster, gestión de conocimiento, algoritmo ISODATA, métodos de agrupación de clúster, sistemas de información, alianzas estratégicas, clasificación documental.

El desarrollo de sistemas de información como consecuencia de la evolución de las tecnologías de información han acaecido conjuntamente con el desarrollo económico y financiero dentro de las organizaciones, lo que inicialmente se planteaba como una herramienta para el desarrollo de funciones operativas, pronto se enmarcó en la base de la gestión empresarial, las comunicaciones y la correcta manipulación de la información entendidas bajo sus principios de confidencialidad, integridad y disponibilidad se convirtió en uno de los principales

motores del progreso empresarial al permitir gestar, gestionar, organizar y comunicar en forma efectiva y eficiente la información relevante y esencial de la organización. Los trabajadores asumieron un nuevo rol como *trabajadores del conocimiento* en donde la autonomía laboral, la libertad en el acceso y manejo de la información, la participación, el talento y el conocimiento tácitoⁱ se convirtieron en nuevas cualidades del rol administrativo y las tecnologías de información debieron enfocar sus desarrollos en la capacidad, la velocidad y el tamaño para gestionar la información.

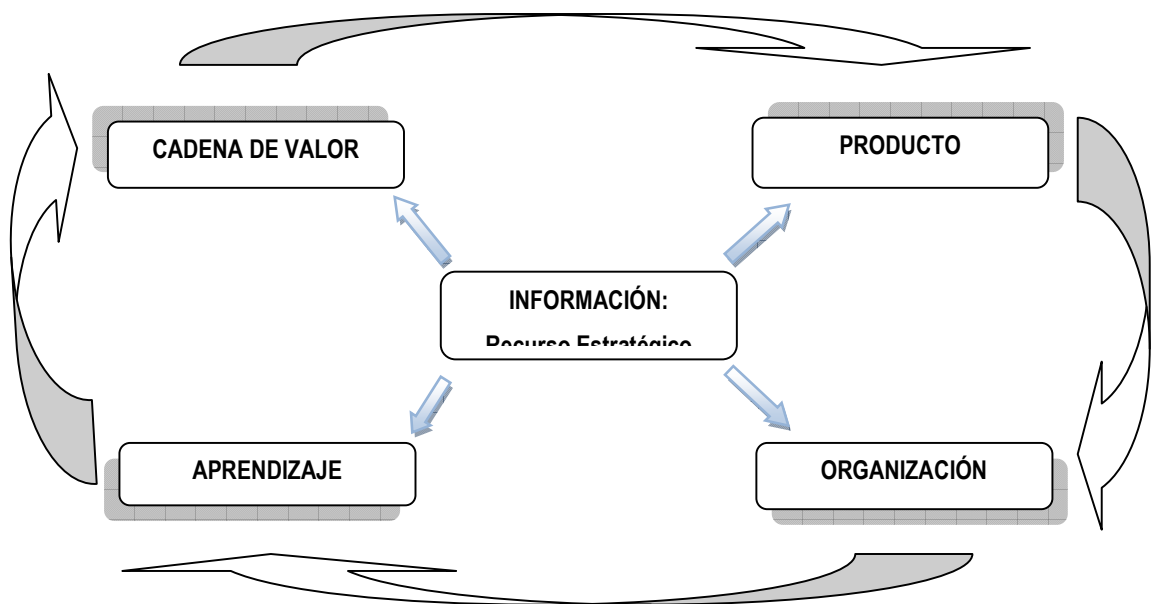


Figura 1. Impacto de la información dentro de la organización.

Las formas de competir a nivel empresarial se dan dentro de la organización a través de la gestión de la información como un recurso estratégico y un activo de la identidad e imagen organizacional, donde la toma de decisiones está supeditada al conocimiento y la gestión de sus fuentes para generar servicios y productos integrales y que obedezcan a las necesidades de sus clientes y los desarrollos de sus proveedores.

Dentro de la gestión del conocimiento y la información organizacional los *clúster* o *clustering*, se han perfilado como una herramienta que permite organizar, acceder y consultar la información de manera eficiente a través de alianzas estratégicas donde se busca fortalecer los puntos débiles de las diferentes áreas organizacionales, brindando estabilidad y simplificando las gestiones administrativas, lo cual desemboca en accesos seguros, veraces y de la más alta calidad con relación a la información de la entidad. Estas sinergias organizacionales promueven la innovación y el desarrollo a nivel interno y externo, involucrando en los sistemas de información empresariales a clientes, proveedores y otros organismos atenuantes para la entidad.

Clustering : Definición, aplicaciones y formas de operación.

El clustering es un método de agrupación mediante el análisis de las palabras que contienen los documentos entre los que existe una relación relevante y que son resultan significativos para las consultas de los mismos. La identificación de los términos dentro de los documentos se puede realizar a través de extractores automáticos o listas de conceptos, los cuales operan fundamentalmente por valor de discriminación y relevancia de términos.

El valor de discriminación permite identificar una parte de un espacio vectorial estadístico para la indización y la recuperación de información, es decir se da un valor más alto a las palabras relevantes y que marcan la diferencia temática de un expediente. El valor por relevancia de términos obedece también al orden sucesivo de la frecuencia de aparición de las palabras obtenidas a partir de los valores de utilidad y de precisión, pero estos son calculados desde algoritmos probabilísticos. Al finalizar la estructura terminológica del expediente, las palabras deben ser organizadas teniendo en cuenta las relaciones semánticas que se dan entre los términos lingüísticos y entre los elementos de los lenguajes documentales, esto permite y asegurar la recuperación efectiva de la información a partir de sus relaciones.

Métodos de Agrupación

Dentro de los métodos de agrupación de clúster más conocidos se encuentran los métodos estadísticos los cuales definen el clúster a partir de un grafo derivado de una media de similitud, esto quiere decir que tomados dos objetos se comparan computando el valor numérico de esta comparación, a modo tal que sobrepasado el coeficiente de similitud por el valor computado son considerados similares; y el método de agrupación basado en masas en donde un término representa a los objetos o términos del clúster, denominado centroide, la similitud de los objetos con el centroide del clúster se mide a través de una función de algoritmos de comparación los cuales utilizan dentro de sus parámetros la identificación del número de clúster deseados, el tamaño mínimo y máximo para cada clúster, el establecimiento del valor para la función de comparación por debajo del cual un objeto no pertenece al clúster, entre otros.

Los algoritmos presentan características muy clara en su funcionamiento, la descripción seriada de objetos, la asignación de objetos a los diferentes clúster a partir de la función de comparación definida y la definición de centroides a partir de los primeros objetos de cada clúster permitiendo la reidentificación de centroides a partir de la evaluación constante de todos los términos.

Otro de los algoritmos más utilizados en los métodos de agrupación, ISODATA *Interactive self organizing data analysis techniques* es un algoritmo interactivo basado en el algoritmo k-vecinos, al que se introduce una consideración heurística que le permite funcionar bien cuando el conocimiento sobre el número de clases no es bueno, los parámetros contenidos por este algoritmo permiten cambiar en forma significativa la características de agrupación del clúster.

Apoyo del desarrollo del conocimiento dentro de la organización: transformación de datos organizacionales en inteligencia empresarial

El objetivo de la clasificación propende la organización documental para su posterior recuperación, cuando los documentos se agrupan en función de ellos

mismos, de su contenido es lo que se conoce como clasificación (automática) no supervisada o clustering; dentro del desarrollo documental empresarial, se puede observar la tendencia a la organización documental por procesos, permitiendo establecer las relaciones y flujos interactuantes entre áreas, es por ello que la cooperación y las alianzas entorno a la investigación, ciencia y desarrollo a nivel tecnológico y organizacional, gestiona la creación de conocimiento dentro de la entidad. Al establecer clustering informacionales se conjugan las relaciones que mantienen las diferentes "dimensiones organizacionales" con algunas capacidades dinámicas como la innovación y el rendimiento financiero y aportar conclusiones relevantes a la hora de analizar la difusión de la innovación, apoyar la toma de decisiones y generar espacios de crecimiento científico y tecnológico. Es importante establecer metodologías existentes para la identificación de clúster informacionales que evidencien el valor de las ventajas de localización, la disponibilidad de recursos, conocimiento e infraestructura informacional y defina las redes de apoyo, las redes inter-organizativas y los problemas de socialización del conocimiento dentro y entre unidades organizativas, como factores determinantes de los procesos de innovación, desarrollo e investigación.

La utilidad de la información dada en la capacidad de almacenamiento, organización, acceso y consulta permite establecer modelos de clúster al interior de la organización para la gestión de la información, obteniendo mejores resultados en la ejecución de recursos, inversiones, toma de decisiones, desarrollo de proyectos y gestión financiera; los clúster dentro de su planteamiento y definición debe tener en cuenta la *relación sistémica* que permite establecer las relaciones entre trabajador, grupo, organización, proceso o función y medioambiente entendido como sociedad; del establecimiento y reconocimiento claro de las relaciones sistémicas nace la estructura del clúster empresarial que se desea manejar para perfeccionar los procesos de dirección como una competencia administrativa que ofrece posibilidades de transformación y cambio. Así mismo la participación activa y democrática en la gestión de los procesos y las

actividades favorecen el surgimiento iniciativas desde lo colectivo.

Los factores base de la estructura del clúster están dados esencialmente por los (trabajadores, proveedores, clientes) individuos y grupos, los procesos en los que participan y métodos y medios que utilizan. El modelo del clúster interacciona con las condiciones organizacionales a partir de la información que posee sobre ellas y que le ofrece “cierto” conocimiento sobre el entorno, permitiendo determinar los objetos, centroides y valores de comparación. Algunos centroides temáticos generales que utilizan las organizaciones para determinar el número de clúster requeridos se encuentran bajo los temas de benchmarking, planes y modelos de gestión, vigilancia económica y tecnológica, inteligencia competitiva, inteligencia económica regional, nacional e internacional y sistemas integrados de información.

BIBLIOGRAFÍA

APPLEGATE, Lyndy M. AUSTIN, Robert D. MACFARLAN, F. Warren. Estrategia y gestión de la información corporativa: Los retos de gestión en una economía en red. 6 ed. Madrid, Bogotá: McGraw-Hill, 2004. 317 p.

CHOO, Chun Wei. La organización inteligente. México: Oxford, 1999. 346 p.

GINER DE LA FUENTE, Fernando. Los sistemas de información en la sociedad del conocimiento. Madrid: ESIC editorial, 2004. 215 p.

HANNA, Nagy. DUCES, Brigitte. BURT, Henry R. A framework for information management. En: Information, technology and facilities department. ([s.l.]: abr, 1950, No.5). p.v.

MAS BASNUEVO, Anays. Modelo para el desarrollo de un sistema de inteligencia

organizacional en la delegación del Ministerio de Ciencia, Tecnología y Medio Ambiente en Holguín. *ACIMED*. [online]. nov.-dic. 2005, vol.13, no.6 [citado 06 de noviembre de 2007]. Disponible en la WWW.: <http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-943520050006000008&lng=pt&nrm=iso>. ISSN 1024-9435.

INFANTE SÁNCHEZ, Diego Esteban. Clustering para recuperación de información. [online] [citado 06 de noviembre de 2007] Disponible en la www.es.geocities.com/clustering_information_retrieval/metodos.html clustering

ⁱ Modelo de conversión del conocimiento desarrollado por Nonaka y Takeuchi (1995).